

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-027985

(43)Date of publication of application : 04.02.1994

(51)Int.Cl.

G10L 3/00

(21)Application number : 04-179704

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>

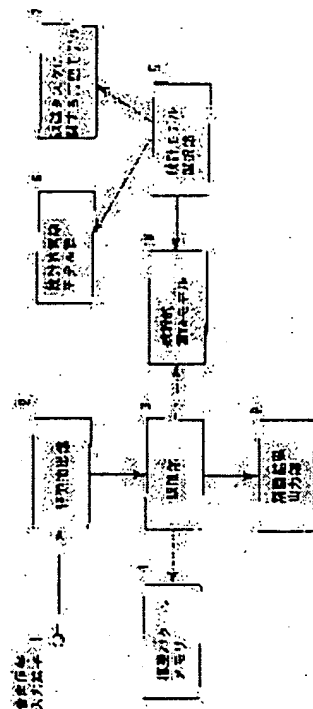
(22)Date of filing : 07.07.1992

(72)Inventor : MATSUNAGA SHOICHI  
KANO KIYOHIRO

## (54) SPEECH RECOGNIZING METHOD

### (57)Abstract:

PURPOSE: To improve the recognizing performance by using a statistical language model which is similar to a vocalization task.  
CONSTITUTION: The statistical language model 6 regarding the occurrence order of syllables, words, etc., is prepared by using respective learning text data bases regarding magazine articles, travel guidancee, conference registration, etc., and a sample of the vocalization task (contents of spoken speech) is used to select a similar statistical language model by a selection part 5; and the selected statistical language model 8 is used to select plural candidate units (syllable and word) with high likelihood whose appearance is predicted and the standard patterns of those selected candidate units are taken out of a memory 4 to find the similarities between those standard patterns and an input speech. Then the sum of the similarities and the occurrence likelihood of the candidate units is regarded as total likelihood and the pattern having the largest 9 likelihood is outputted as a recognition result.



## LEGAL STATUS

[Date of request for examination]

07.01.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3240691

[Date of registration]

19.10.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(11)特許出願公開番号

特開平6-27985

(43)公開日 平成6年(1994)2月4日

(51)Int.Cl.<sup>5</sup>  
G 1 0 L 3/00

識別記号 庁内整理番号  
531 D 7627-5H

F I

### 技術表示箇所

審査請求 未請求 請求項の数 1 (全 5 頁)

(21)出願番号 特願平4-179704

(22)出願日 平成4年(1992)7月7日

(71)出願人 000004226  
日本電信電話株式会社  
東京都千代田区内幸町一丁目1番6号

(72)発明者 松永 昭一  
東京都千代田区内幸町1丁目1番6号 日  
本電信電話株式会社内

(72)発明者 鹿野 清宏  
東京都千代田区内幸町1丁目1番6号 日  
本電信電話株式会社内

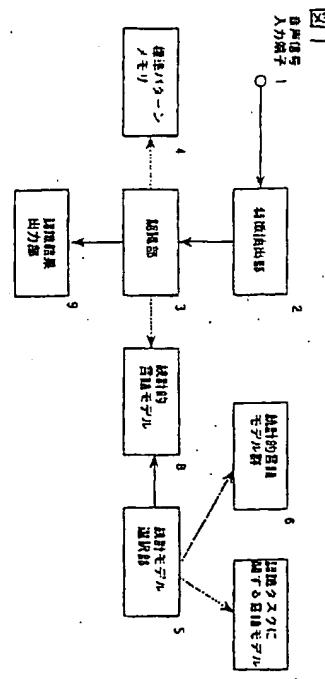
(74)代理人 弁理士 草野 卓

(54)【発明の名称】 音声認識方法

(57) 【要約】

【目的】 認識性能を向上させる。

【構成】 雑誌記事、旅行案内、会議登録などに関する各学習用テキストデータベースからそれぞれ音節、単語などの生起順序に関する統計的言語モデル6を予め用意しておき、発声タスク（発声音声の内容）のサンプルを用いてこれと類似した統計的言語モデルを選択部5で選択し、その選択した統計的言語モデル8を用いて、それまでの認識結果から次に出現されると予測される尤度の高い複数の候補ユニット（音節や単語）を選出し、これら選出した候補ユニットの標準パターンをメモリ4から取出し、これら各標準パターンと入力音声との類似度を求め、これら類似度とその候補ユニットの生起尤度との和を総合尤度とし、これが最も大きいものを認識結果として出力する。



## 【特許請求の範囲】

【請求項1】 入力音声の特徴パラメータの時系列とし、生起順序に関する統計的言語モデルを用いて、上記入力音声の特徴パラメータ時系列について、複数の音声認識候補を選出し、これらの各音声認識候補について、音声標準パターンと上記入力音声の特徴パラメータ時系列とをそれぞれ照合して、生起の尤度と類似の尤度との総合尤度の高い候補を認識結果とする音声認識方法において、

異なるタスクの学習用テキストデータベースからそれぞれ作成された生起順序に関する統計的言語モデル群を予め用意しておき、

発声するタスクのサンプルを用いて、発声タスク（発声する音声の内容）と類似した統計的言語モデルを上記統計的モデル群から選択し、

その選択した統計的言語モデルを、上記複数の音声認識候補の選出に使用することを特徴とする音声認識方法。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 この発明は、統計的言語モデル（例えば、Bahl, L. 他 "A. Statistical Approach to Continuous Speech Recognition" IEEE Trans. on PAMI (1983)）を用いた音声認識方法に関する。

## 【0002】

【従来の技術】 従来の統計的言語モデルを用いた音声認識方法として、学習用テキストデータベースより、音節や単語の生起順序に関する統計的言語モデルと、音節や単語の標準パターン（例えば、隠れマルコフモデル等）とを予め作成しておき、入力音声に対し、統計的言語モデルを用いて、既に認識した直前の複数の音節や単語から、次に生起する確率の高い複数の音節や単語候補を選 \*

$$D(A, B) = (\sum (P_n(A) - P_n(B)) (\log P_n(A) - \log P_n(B))) / N$$

で表す。ここで、 $\sum$ は $n=1$ から $N$ まで、 $N$ は要素数（例えば、三つ組の種類数）である。 $D$ の値が小さい程、 $A, B$ は類似している。

## 【0006】

【実施例】 図1に、この発明の実施例を示す。入力端子1から入力された音声は、特徴抽出部2においてデジタル信号に変換され、更にLPCケプストラム分析された後、1フレーム（例えば10ミリ秒）ごとに特徴パラメータに変換される。この特徴パラメータは例えばLPCケプストラム係数である。

【0007】 予め、学習用音声データベースより、上記特徴パラメータと同一形式で、音節あるいは単語の標準パターンを作り、標準パターンメモリ4に記憶しておく。また、統計モデル選択部は5、認識に用いる統計的言語モデル8を、統計的言語モデル群6の統計的言語モ

\*出し、これら選出した音節や単語候補のそれぞれについてその音節や単語の標準パターンと入力音声とを照合して、生起尤度と標準パターンとの類似尤度との総合尤度の最も高い音節や単語を認識結果として出力することが提案されている。

## 【0003】

【発明が解決しようとする課題】 しかし、この認識方法は統計的言語モデルが、認識タスク（発声内容）に類似した大規模なものでなければならなかった。例えば、統計的言語モデルが新聞の社説の大規模なデータベースから作成されたものであれば、社説の内容の発声に対する認識に有効であるが、例えば電話会議登録などの異なるタスクに対する認識には有効性が低かった。つまり、異なるタスクを認識させる場合にはそれに応じた大規模なテキストデータを作成しなければならなかった。

## 【0004】

【課題を解決するための手段】 この発明によれば、異なるタスクのテキストデータベースからそれぞれ生起順序に関する統計的言語モデル群を予め用意しておき、その統計的言語モデル群の中から、予め得た少量の発声用テキストと類似した統計的言語モデルを自動的に選出し、この選出したモデルを音声認識用統計的言語モデルに用いる。このように適応型統計的言語モデルを用いて認識することで、任意の統計的言語モデルを用いる場合より認識性能が上昇する。

【0005】 統計的言語モデルの選択の手法としては、例えばKullbackのdivergence（坂井利之他「パターン認識の理論」共立出版（1967））を用いればよい。具体的には、2つの統計的言語モデルを $A, B$ で表し、モデル間の距離を $D$ 、モデルの各要素（例えば、文字の三つ組（例えば、「あい」等）の出現する確率、トライグラム）をモデル $A$ に関して $P_n(A)$ 、モデル $B$ に関して $P_n(B)$ 、とすると、

デル（ $M_1, M_2, \dots, M_I, I$ は言語モデルの数）の中から、認識タスクに関する言語モデル7に最も類似している統計的言語モデル8（ $M_j$ ）を選択する。新聞の社説、会議登録、旅行案内など異なるタスク（発声内容）についての各学習用テキストデータベースからそれぞれの音節や単語の生起順序に関する統計的言語モデルを予め作成しておき、これらを統計的言語モデル群6とする。

【0008】 音声認識部3では、選択した統計的言語モデル8（ $M_j$ ）を用いて選出した複数の音節や単語の候補について、その候補の標準パターンを標準パターンメモリ4から読みだし、入力音声のパラメータとの類似度（尤度）をそれぞれ求める。つまり例えば入力音声の $i$ 番目のユニット（音節や単語）を認識するには、選択した統計的言語モデル8からユニットの出現順序に関する

トライグラムを用いて、 $(i-2)$  番目と  $(i-1)$  番目との各ユニットの認識結果を基に、 $i$  番目に出現されると予測される尤度が高い複数のユニットを候補ユニット  $k_1 \sim k_n$  として選出する(図2)。これら選出された各候補ユニット  $k_1 \sim k_n$  に対する標準パターンと入力音声との尤度(類似度)をそれぞれ求め、その候補ユニットの  $i$  番目に出現する尤度と、その標準パターンとの類似性を示す尤度との和を総合尤度とし、この尤度が最も高い候補ユニット、例えば  $k_2$  を  $i$  番目の認識結果として認識結果出力部9へ出力する。

【0009】このユニット候補の選出と、それらについての標準パターンとの照合と、その総合尤度から認識結果ユニットを得る操作とを音声区間が終わるまで繰り返し、最後に、それまで得られた認識結果ユニットを、その順に入力音声の認識結果の系列として出力する。なお、特徴抽出部2、認識部3、認識結果出力部9、統計的言語モデルの選択部5はそれぞれ専用、または兼用のマイクロプロセッサにより処理することができる。

【0010】更に選択部5はKullbackのdivergenceに限るわけではない。たとえば、ユークリッド距離等の距離が算出できるいずれの尺度でもよい。また統計的言語モデルや認識標準パターンのユニットは音節や単語だけでなく、音素やかな、漢字などの文字単位であってもよい。認識手法は隠れマルコフモデルに限らず、DPマッチングを用いても良い。統計的言語モデルもトライグラムに限らず、バイグラムやユニグラムの統計量でも良い。

【0011】

【発明の効果】以上述べたように、この発明によれば、発声タスクと類似した統計的言語モデルを用いるため、

任意の統計的言語モデルを用いる場合よりも高い認識性能が予期される。会議登録に関する発声タスク279文節に対して文節認識率による評価を行った。例えば、旅行案内に関するテキストを統計的言語モデルとして用いた場合、認識性能は42%であるが、会議登録に関するテキストを用いた場合には認識性能は64%に上昇した。

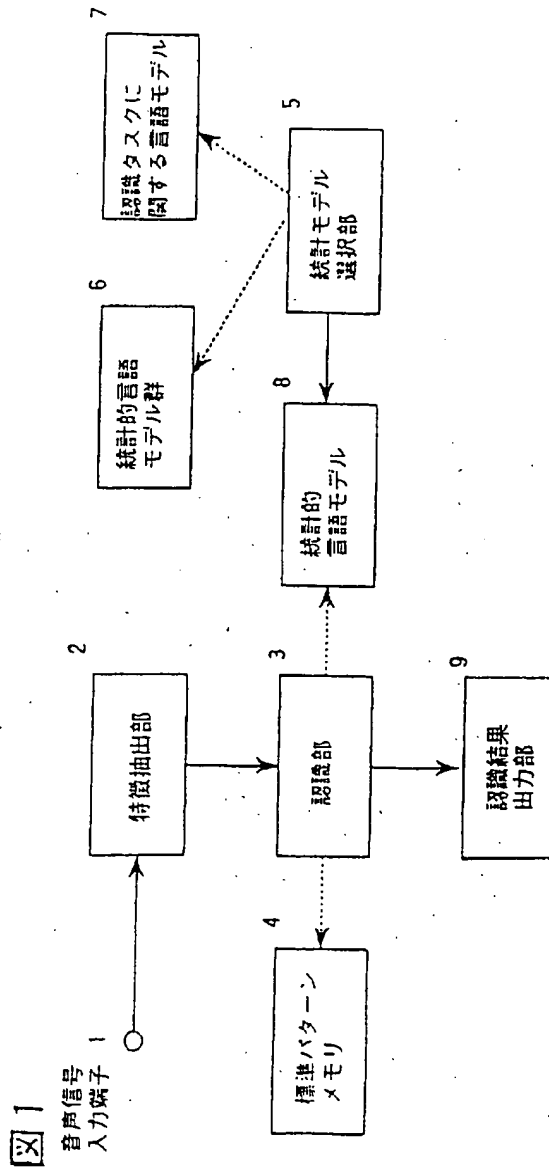
【0012】これに対して、雑誌記事、論説文、新聞、会議登録に関するキーボード会話の4種の統計的言語モデルの中から会議登録に関する電話会議と最も類似した統計的言語モデルを、この発明方法の中の統計モデル選択部を用いて選んだ場合、会議登録に関するキーボード会話が選ばれ、本選択が適切であることが示された。なお、複数の統計的言語モデルの群を用意しておくことなく、雑誌記事、論説文、新聞、旅行案内、会議登録などの各学習用テキストデータベースを用意しておき、発声タスクのサンプルを用いて、これと類似した学習用テキストデータベースを選択し、その選択した学習用テキストデータベースから生起順序に関する統計的言語モデルを作り、この統計的言語モデルを音声認識に使用してもよい。しかしこの場合は各種の学習用テキストデータベースを用意しておくために、より多くの記憶容量が必要であり、かつ発声タスクと学習用テキストデータベースとの類似性を調べるには統計量を用いることになり、統計的言語モデルとの類似性を調べる方が処理が簡単となる。

【図面の簡単な説明】

【図1】この発明の実施例を示すブロック図。

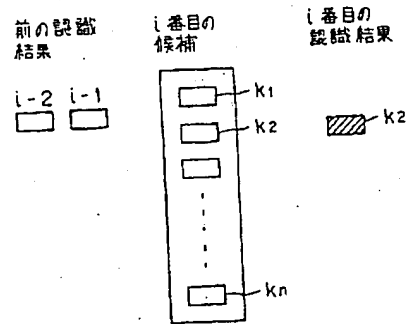
【図2】 $i$  番目の認識のための候補ユニットを選出し、これより認識結果を出力する説明図。

【図 1】



【図2】

図 2



□ はユニット (音節や単語等) を示す。